607
Applied work as data reduction

Jon Faust

http://e105.org/e607

September 5, 2016

▶ **Readings**

- There is no specific reading for this riff
- But now would be a good time to begin reading through the articles in the history section of the reading list.

▶ **What is applied econometrics**

▶ **Applied econometrics is...**

- In applied econometrics, we start with 'raw' data
- And transform it in some way that folks find useful.
- For example, we highlight features that people find useful or interesting

  Or perhaps we highlight things and attempt to convince folks that these are important or interesting.

▶ **Applied econometrics in a rational expectations world**

- In rational expectations equilibrium, agents understand all implications of public information
- In an RE world, we would just create new datasets and then we'd post them on the Web.
- All implications of the public data would then be incorporated in the equilbrium.

▶ **Applied econometrics in practice**

- In practice, you can make a very good living transforming raw data into some much lower dimensional set of summary statistics

  Point estimates of a parameterized model, test statistics, summaries of posterior distributions, etc.

- In doing so, you are highlighting implications of the data that folks had not previously noticed and find useful in some way.

▶ **Applied work as data reduction**

- Call the totality of all of all possibly relevant data, W.
- In empirical work, we reduce or compress W into some much, much, lower dimensional set of things to report.

  A few statistics and figures.

- The essence here is stripping away the less important part of the data (for the purpose at hand) and leaving some part that is useful.

  This is sometimes called data reduction

▶ **Kind of formally**

- Empirical work is mapping, say, γ from W to a much lower dimensional V:

$$\gamma : \mathcal{W} \to V$$

  where V is much lower dimensional than W.
- We get a particular dataset, W^r and report $v = \gamma(W^r)$.

▶ **No loss of information**

- In a formal framework, we can ask how far we can reduce things and literally lose nothing of any value.
- The first key notion here is a sufficient statistic.

▶ **Sufficient statistic**

- Suppose $W \sim P_\theta$ and that the only unknown in the problem is θ.
- In words, $\gamma(W)$ is a sufficient statistic for θ iff $\gamma(W)$ is as informative about θ as is W.

  Any decisionmaker would be indifferent between observing W and $\gamma(W)$.

- As you should know, if $w_i \sim iidN(\mu, \sigma^2)$, $t = 1, \ldots, T$, then $\gamma(W) = \{\sum w_i \text{ and } \sum w_i^2\}$ is sufficient for $\theta = \{\mu, \sigma^2\}$.
- Thus, we can compress an arbitrarily large dataset into two 2 numbers.
- Fomally, if $W \sim P_\theta$, γ is sufficient for θ iff we can factor P as,

$$P_\theta(W) = Q(\gamma(W), \theta) \times \phi(W)$$

- That is, the probabilty distribution can be factored into a piece depending on the data only through γ and a piece that does not involve θ.
- You should be able to verify that, for example, the MLE of θ can be written only in terms of sufficient statistics $\gamma(W)$.

▶ **Additional senses of loss-less or low-loss reduction**

- If you dig deeper into theory you'll learn about various senses in which problems admit lossless reductions or reductions that are not important from some perspective.

  Invariant problems and invariant statistics, various forms of ancillarity, and so forth.

- See Lehman, Testing Statistical Hypotheses for an excellent introduction to these concepts.

  This is a very beautiful book by one of the greats...

▶ **Decision theory**

- We could come at reduction a slightly different way by defining a decision problem and a loss function over mistaken decisions
- And we could think of all possible ways to map data into a decision.
- Then we could ask when a set of decision rules big enough to include all reasonable approaches
- One concept here is a *complete class* of decision procedures.

▶ **x**

- Suppose $\mathcal{D}$ is the set of all decision rules and $\mathcal{C} \subset \mathcal{D}$.
- $\mathcal{C}$ forms an essentially complete class if for every decision rule in $\mathcal{D}$ there is a rule that attains at least as low a loss in $\mathcal{C}$.
- In short, we are not made worse off by limiting ourselves to decision rules in $\mathcal{C}$.
- One reason to take this digression is that the set of Bayesian decision rules (and limits of sequences of such rules) generally form a complete class.

  See Lehmann once again.

- In this sense, even if you are a frequentist, you know that you can limit yourself to Bayesian decision rules without loss.
- This is loose talk, which you may want to learn better, but I won't drag you through the details.
- I mainly state this last result to further illuminate the inherent ties between Bayesian and frequentist approaches.

▶ **Reality crashes in**

- You won't hear a lot more about loss-less reduction in this course
- As it turns out, our macro problems are often messy enough that there is no low-dimensional summary of the data that we can be confident invovles little loss of relevant information.
- Thus, we can either report a set of statistics as large as the orginial dataset, or we can report a much smaller set of statistics and face the fact that we have probably omitted relevant information.
- Thus, the art of choosing wisely is an important part of applied econometrics.

▶ **Aside:: Questionable work**

- These days one can find papers with pages and pages of statistics.
- A careful count might reveal that the papers include more statistics than data points in some cases.
- It's usually best to avoid such work.

▶ **What makes macro problems messy?**

▶ **What makes macro problems messy?**

- 1. Subject matter is aggregates, and the assumptions of aggregation theorems do not hold even approximately in practice.

  Thus, can't naively reason from individuals to aggregates

- 2. General equilibrium effects are often important

  in GE, loosely speaking, everything affects everything

- 3. Expecations effects often important

  anticipations of everything also affect everything

▶ **Adding up these three**

- Lemma: Almost nothing can be plausibly taken to be exogenous
- Lemma: There are almost no valid instruments for any macro phenomena

  This second lemma is arguably a main point of Sims famous article (see the reading list) Macroeconomics and Reality. You should all read this.

▶ **Why is applied macro hard**

- 4. Standard samples have 20-50 years of data, which is a short span relative to the duration of studied phenomena.

  Recession every 10 years; crisis every 50; variables show 20 and 30 year secular trends

- Even when we are studying more short-lived phenomena, the possible general equilibrium interacations with longer-run phenomena may confound our results.
- 5. 'fresh data' arrive slowly

  We seldom discover/create a new, relevant dataset and so fresh data arrive at the rate of approximately one annual observation per year.

- Possible exceptions
  - Create new historical data
  - Examinining new data for other countries than those previously studied.
- Relevance of data under these exceptions is often questionable

  Economic institutions and structures change

- 6. Macro data are not from controlled experiments
- Thus, we know little about underlying structure

  about distributions of shocks, etc.

- 7. Macro theory as it has evolved so far does not give precise predictions on many aspects of macro systems
- For example, dynamics are a key issue, but modest changes in assumptions about small 'frictions' often lead to important changes in dynamic implications

► **Overall,**

- We have datasets with lots of variables but for short time spans (relative to phenomena of interest). Theory says they may be richly related at a wide variety of leads and lags, but theory does not nail down dynamics very specifically.
- We dramatically reduce the full macro dataset in most applied papers

  e.g., 1 to 10 conventional data series for a short sample

- And we must face the possibility that we have omitted relevant info.
- Drawing reliable conclusion from macro data is difficult
- This is good (so long as you are humble in your ambitions)
- Because you can learn how to do this better than others and this scarce skill will be rewarded

► **Wrapping up**

- **David**

- **Michaelangelo**
  - Michaelanglo was reportedly asked how he created the sculpture of David.
  - And he responded, 'take the block of stone and chip away the part that isn't David.'
- **Reduction: Applied papers are an exercise in taking a full dataset and** chipping away less informative stuff and leaving something that is informative about some question of interest.
  - Sadly, we always have to chip away a lot of stuff that may be of interest

    Say, David's head, some of the torso and one leg, and both arms.
  - And what we leave may not be as beautiful as we'd like.

- But we can still aspire to elegance or at least usefulness.
- And we can still aspire to scientific standards, which should lead over time to progress.